DOCUMENT RESUME

ED 161 944 TM 007 979

AUTHOR James, Barry N.
TITLE An Examination of an Approach for Using Pupil Outcome Scores to Validate Assessments of Teacher Behaviors.
PUB DATE Mar 78
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)
AVAILABLE FROM Teacher Education Projects, 318 Johnston Building, 415 North Monroe Street, Tallahassee, Florida 32304 ($2.00 ea.; 10 or more copies, $1.80 ea.)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Academic Achievement; *Classroom Observation Techniques; *Effective Teaching; Evaluation Criteria; Evaluation Methods; Grade 3; Higher Education; *Performance Based Teacher Education; Predictive Validity; Primary Education; Student Teachers; Teacher Behavior; Teacher Evaluation; *Teaching Models; *Teaching Skills; Test Validity

ABSTRACT

A system which used pupil outcomes to estimate the predictive validity of assessments of teacher outcomes was field tested in a performance-based teacher education program at Florida State University. Three levels of teaching competency were assessed: (1) competencies associated with cognitive identification of appropriate decisions for implementing a teaching model; (2) competencies for performing model-related behaviors; and (3) competencies associated with inducing pupils to demonstrate model-related classroom behaviors. Five instruments, measuring teacher knowledge and classroom performance and student classroom behavior and outcomes, were used to evaluate program effectiveness and student teachers' readiness to teach in public schools. The field test was conducted in a public school setting with 24 teacher candidates and 120 third grade students. Each candidate delivered six half-hour lessons to five students during a four-week period. In addition to the measurement instruments, videotapes were used by three-person observation teams to assess teacher classroom performance and student classroom performance. Correlation coefficients and multiple regression analysis indicated that pupil outcome scores may be successfully used to estimate the predictive validity of assessment instruments and to aid in the validation of teaching competencies within the teaching model framework. (Author/JAC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED161944

An Examination of an Approach for Using Pupil Outcome Scores to Validate Assessments of Teacher Behaviors

the national conference of the American Educational Research Association, Toronto, 1978

by Barry N. James

Florida State University/Florida A & M University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Barry N. James

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

In a performance-based teacher education program at Florida State University, a system was field tested which involved attempts to use pupil outcomes to estimate predictive validity of assessments of teacher competencies. Results indicate that predictive validity may be established if teaching competencies and pupil outcomes are assessed within a models of teaching framework.

As implied in the title, this paper is concerned with attempts to demonstrate relationships between teacher behaviors and pupil performance. Specifically it reports results of an initial trial of an approach for using pupil outcome scores to estimate *predictive validity* of assessments of teacher behaviors. The results also have implications for attempts to validate teaching competencies. However, the study to be reported was only one facet of a broader field test of an assessment system for a performance-based teacher education (PBTE) program, and to provide a basis for evaluating results, the paper begins with a brief description of problems associated with assessment in PBTE and of the system evaluated.

ASSESSMENT IN PBTE PROGRAMS

Although a wide variety of PBTE programs have been developed, nearly all are characterized by the following three elements (Cooper, Weber, and Johnson, 1973):

1. Specification of competencies to be demonstrated by each teacher candidate.
2. Specification of explicit criteria for assessing competencies of teacher candidates.
3. A requirement that each teacher candidate demonstrate specified competencies.

Assessment plays an important role in PBTE for at least three reasons. Obviously, it is through assessment activities that competency acquisition is demonstrated, and assessment is a source of information needed for program improvement. Thirdly, assessment activities must provide the data that probably is necessary to the survival of the movement—evidence that graduates of PBTE programs are more effective teachers than graduates of other programs (McDonald, 1972). Unfortunately, not only are assessment instruments for PBTE programs difficult to develop and validate it has also been difficult for proponents of PBTE to demonstrate that teachers with any identified sets of competencies are better teachers than teachers without those competencies. In other words, little direct relationship has been found between teacher behaviors and pupil outcomes. Until that relationship is established, it may not be possible to demonstrate the effectiveness of competencies addressed in a PBTE program, to validate the procedures used to assess those competencies, and to evaluate training programs designed to produce or nurture those competencies.

The primary obstacle to competency validation has been a general absence in teacher behavior-pupil performance research of statistically significant results (Berliner, 1975; James, 1977; Heath and Nielson, 1974; Potter, 1973; Rosenshine and Furst, 1971). Most of the studies have addressed effects of singular teacher skills or behaviors, but studies describing classroom processes at a molar level, or involving clusters of skills, have been more successful (Berliner, 1975; Medley, Soar, and Soar, 1975). These results indicate that it may be inappropriate to measure skills in isolation. Even the set of indicators used to demonstrate that a competency has been acquired is influenced by the expected outcome and other competencies associated with that outcome. For example, if a candidate is to be assessed on his/her ability to maintain large-group control, what indicators would imply that the candidate was competent? The answer to that question may be dependent on the expected outcome. Large-group control conducive to verbal information learning may be very inappropriate in lessons related to information processing or social skills. Maintenance of large-group control may even be a skill generic to all outcomes, but the characteristics of the application of that skill would differ across outcomes. A teacher must maintain appropriate large-group control while simultaneously demonstrating other competencies, all of which should result in an expected outcome.

TM007 979

© 1978 by Barry N. James and Florida State University Foundation

Another common characteristic of teacher behavior-pupil performance research has been the administration of achievement tests which measure a wide range of pupil skills. The use of such tests to measure the effects of a competency, or set of competencies, implies that a single competency may be expected to positively influence several, if not all, domains of pupil outcomes. There may be a few competencies with universal teaching application, but there is little reason to believe that most competencies will positively influence all outcomes, or even all the outcomes measured in any standardized test. The fact that a competency does not influence all outcomes may not reduce its importance, for even if the competency is essential to only one outcome, the competency may be worth acquiring if the outcome is important. Therefore, it seems inappropriat to attempt to decide whether or not a competency is valid on the basis of scores on a standardized test that measures a variety of outcomes. Even if universal application were a reasonable expectation for a competency, or a set of competencies, available tests may provide questionable measures of pupil achievement. Several researchers (Anastasi, 1968; Ekstrom, 1973; Gagne, 1970; Gall, 1973) have noted the similarity of most achievement tests to intelligence tests. Anastasi (1968) found that scores on intelligence tests correlate as highly with achievement tests as with other intelligence tests.

A premise examined in this study is that relationships between teacher performance and pupil outcomes can be demonstrated if the following two conditions are met.

1. Clusters of competencies which are related to specific, designated pupil outcomes are examined as a unit rather than as isolated, unrelated competencies.

2. A set of competencies is examined in terms of its relationship to specific, designated pupil outcomes rather than a universal set of academic skills.

A means of satisfying these assumptions may be found in the work of Joyce and Weil (1972).

MODELS OF TEACHING

Joyce and Weil (1972) have examined and described nineteen interactive models of teaching based on various learning theories and have suggested that no model is always appropriate. The decision as to which model should be applied in a classroom situation is dependent on the desired pupil outcomes and the characteristics of involved pupils. A model of teaching is "a complex of behavioral events in which a teacher carries out a sequence of activities designed to implement particular educational objectives and goals" (Hunt, Joyce, Greenwood, Noy, Reid, and Weil, 1974, p. 70).

Each model has been described in terms of the following five concepts.

1. The orientation or focus of the model: a description of the goals, educational theory, and important aspects of the pupils' environment.

2. The syntax of the model: a description, in phases, of teacher interactive behavior in the classroom.

3. The principles of reaction of the model: a description of a teacher's reactions to pupil behaviors.

4. The social system characteristic of the model: teacher and pupil roles, authority relationships, and behavior norms which are rewarded.

5. The support system required by the model: a description of any special environmental conditions, such as equipment and facilities, or personal skills (e.g., extensive interactive skill) essential to the success of the model.

Dodl and Freedman (1974) have added the concepts of preactive and postactive support. Preactive support refers to those activities—such as selection of materials, planning of lessons, and designing of evaluation instruments—which must be performed prior to interactive classroom instruction. Postactive support involves activities, such as evaluation data collection and analysis, which are performed after interactive classroom instruction and which are essential to preparation for future instruction.

MODELS OF TEACHING AS A FRAMEWORK FOR ASSESSING TEACHER COMPETENCIES

Dodl and Freedman (1974) have developed—and Freedman (1975) has validated—a procedure for deriving competencies which would be required by teachers attempting to apply a model of teaching. Utilization of a list of competencies derived from a model and use of instruments designed to assess pupil skills/values which that model is intended to produce or nurture could satisfy conditions listed earlier for examining the relationship between teacher competencies:

1. Clusters of competencies (derived from a single model) can be examined as a unit rather than as isolated, unrelated competencies.

2. A set of competencies can be examined in terms of its relationship to specific, designated pupil outcomes (also derived from a single model) rather than to a universal set of academic skills.

A RESEARCH PARADIGM

Medley, Soar, and Soar (1975) have been studying assessment problems related to PBTE programs for several years, and have suggested that the paradigm graphically summarized in Figure 1 would appear to be appropriate for examining teacher competency.

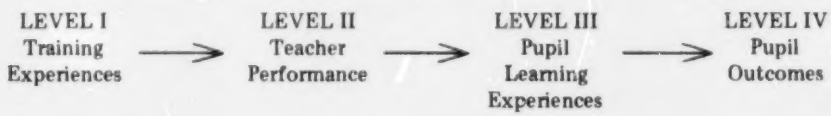

Figure 1: Paradigm for assessment research

Level I refers to assessment of the training experiences a candidate might have had before teaching pupils in classrooms. These assessments might be nothing more than listings or descriptions of completed coursework, lists of completed modules, scores on comprenensive tests, or any information depicting how the candidate spent his training time prior to teaching.

Level II assessments are based on observations of the behaviors of a candidate as she/he performs in the role of a teacher.

Level III involves observations of the experiences of pupils, in and out of class, while the candidate is filling the role of a teacher.

Level IV relates to pupil performance outcomes which are representative of the intended effects of teaching.

The arrows in the paradigm suggest that each level has a cause-and-effect relationship to the other levels. For instance, pupil outcomes (Level IV) are assumed to be directly related to pupil learning experiences (Level III) which are related to teacher performance (Level II) and training experiences (Level I). Eventually, assessments at any of the first three levels must be validated by assessments at Level IV. A teacher training program is valid only if pupils learn as a result of the skills or knowledge attained by teachers. Likewise, what a teacher does in the classroom or what learning experiences pupils have are important only if they eventually affect pupil outcomes. When discussing competencies, Medley, Soar, and Soar equate effectiveness and validity. In other words, valid competencies are those which produce or nurture pupil outcomes.

It would appear that Level IV assessments would provide the most appropriate measures of teacher competencies, but Medley, Soar, and Soar suggest that Level IV assessments are probably inappropriate tools for determining if a teacher is ready to teach (i.e., to be certified). However, it should be noted that the study reported in this paper is not concerned with determination of a teacher candidate's qualifications for certification, but rather the validity of competencies candidates are to demonstrate. It is suggested that assessment systems employed in PBTE programs should have two distinct functions: (a) assessment of candidate teaching competence, and (b) research concerning the effectiveness of competencies, assessment procedures, and

training programs. It is recommended that candidates be certified on the basis of Level II (teacher classroom performance) assessments (and in exceptional instances to be discussed later, Level I assessments), but during the assessment process, simultaneous attempts should be made to validate assessments in terms of pupil outcomes. Before and after competencies are validated, candidates should be judged according to skills they demonstrate in classrooms, not on their effect on pupil outcomes, except perhaps on short-term outcomes such as those typically addressed in minicourses (Borg, Kelley, and Gall, 1970).

Medley, Soar, and Soar (1975) also suggest that assessments at Levels III and IV have questionable validity, reliability, and practicality; but their concerns are addressed when the paradigm is employed within a models of teaching framework. A detailed discussion of those questions can be found in the initial evaluation report on the entire assessment system (James, 1977).

REVISION OF THE ASSESSMENT PARADIGM

With the following revisions, the assessment paradigm described by Medley, Soar, and Soar was utilized for this study within a models of teaching framework.

1. Level I assessments, in the present study, refer to measures of a candidate's knowledge about a model. Paper and pen tests are used at this level, and items are based on decisions which a teacher must make when designing and performing a model. Level I assessments are usually discouraged in PBTE programs, but it may be found that, with some models, a candidate who can best make model-related decisions can best perform the model in the classroom and best generate appropriate pupil classroom behaviors and pupil outcomes. It may further be found that Level I tests consistently correlate with scores of assessments at other levels. If so, a Level I assessment could measure those decision-related skills and provide an exceptionally *practical* measure of teacher candidate preparedness. If Level I scores consistently correlate with Level IV scores, Level I scores would have value as predictors of effectiveness with pupils. At least the assessment instrument could be utilized to predict a candidate's readiness to participate in a Level II assessment.

2. Level II assessments, in the present study, refer to measures of a candidate's classroom performance. Competencies assessed are associated with the syntax and principles of reaction of a model of teaching. Observation instruments were designed to determine if teachers were correctly performing instructional activities and were responding appropriately to pupil behavior, according to the model.

3. Level III assessments refer to measures of a candidate's ability to promote appropriate pupil classroom behaviors. Competencies to be assessed are associated with the social system related to the model and pupil behaviors implied in the syntax of the model.

4. Level IV assessments refer to a candidate's ability to teach pupils to demonstrate the skills associated with the goals of a specific model. Pupils taught by a candidate were asked to perform tasks which required skills related to model outcomes.

In this study, assessment levels and competencies were considered to be related; or restated, three levels of teaching competency were assessed in the study:

1. Level I: competencies associated with cognitive identification of appropriate decisions for implementing a model.

2. Level II: competencies associated with performing model-related behaviors.

3. Level III: competencies associated with inducing pupils to demonstrate model-related classroom behaviors.

The revised paradigm is illustrated on the top of page 5.

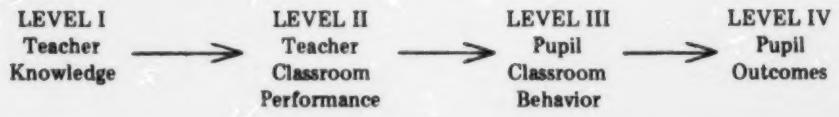

Figure 2: Revised paradigm for assessment research

THE FIELD TEST

An adequate system within a PBTE program must not only facilitate assessment of competencies possessed by teacher candidates, but must provide data which can be utilized to answer the following questions (Burke, Hansen, Houston, and Johnson, 1975).

1. Are competencies effective?
2. Are assessment instruments effective?
3. Are training procedures effective?
4. What are the most practical assessment procedures?

In the system examined in this study, only those competencies and expected pupil outcomes associated with a specific model are assessed. Teacher candidates in an experimental group are trained to use a specific model and are asked to teach to pupils the skills, knowledge, or values which the model is designed to produce or nurture. Candidates in a control group are asked to try to achieve the same pupil outcomes, but candidates are taught to use teaching methods selected by the instructor of the control group. However, when data was analyzed for the purpose of addressing competency validation, group scores were collapsed. Such analysis was concerned with the effects of competencies, not which group had acquired the competencies.

In the field test, the system was evaluated in terms of three questions:

1. Was the necessary data yielded for the system to serve the functions for which it was designed?

2. What additional impact did the system have on the operations of the public school or teacher training programs in which it was implemented?

3. What additional costs (in money and time) were necessary to implement the system?

However, in this paper only those results directly related to validation of assessment procedures are discussed, but it can be reported that results were positive (James, 1977).

The field test was conducted in a public school setting in which twenty-four teacher candidates were participating in a PBTE program. The experimental group consisted of twelve candidates who were taught to use the Concept Attainment Model associated with the work of Jerome S. Bruner (Joyce and Weil, 1973), and the twelve candidates in the control group were taught to use a problem-solving model selected by the instructor of the control group. After five weeks of training, all candidates were given a Level I (knowledge) multiple-choice assessment, and 120 third grade pupils were given a Level IV (pupil outcomes) paper and pen test (on which no pupils could perform any tasks). Then, five of the third grade pupils were randomly assigned to each candidate, who was asked to achieve the following goal with his/her assigned pupils:

> The pupils will be able to attain a concept by examining the attributes of examples and non-examples of the concept.

Each candidate delivered six half-hour lessons to pupils during a four-week period, and during the first lesson, the candidate and his/her pupils were videotaped. At the end of the training period, pupils were given a Level IV posttest, and two three-person observation teams used the videotapes to compute scores for assessments at Level II (teacher classroom performance) and Level III (pupil classroom behaviors).

ASSESSMENT INSTRUMENTS

Instruments were designed to accumulate four types of scores for each teacher candidate:

1. The number of correct responses on a cognitive-level multiple-choice test given to all candidates.
2. The number of model-related competencies demonstrated in a videotaped lesson.
3. The number of model-related pupil behaviors demonstrated in a videotaped lesson.
4. The mean score of pupils on a model-related pupil-outcomes test.

UTILITY OF INSTRUMENTS

The instruments used in the proposed assessment system should not only generate data for evaluating program effectiveness, but also measure each candidate's readiness to teach in public schools. It is unlikely that the instruments would be utilized for the latter function unless they were reasonably easy to administer. Although it might be possible to design a low-inference Level II instrument that would enable an evaluator to obtain frequency counts of discrete teacher behaviors (as advocated by Medley, Soar, and Soar, 1975), a set of teaching-model-related competencies could require numerous frequency counts, and usually only a few frequency counts can be taken in one viewing. An evaluator might be required to view one videotape numerous times, and it is doubtful that teacher training personnel have sufficient time to view each intern's tape several times each time the intern needs to demonstrate her/his ability to apply a model.

In the interest of utility, therefore, it was decided that observation instruments would have a fairly high inference level. In addition, the number of items included in instruments was limited in an effort to reduce the time required to administer instruments.

Two faculty members (including one of the authors of the Competency Derivation Procedure) experienced in teaching models of teaching determined face validity for all instruments. In addition, a graduate class which was currently studying the Concept Attainment Model was asked to field test and criticize the Level I test, and the test was revised on the basis of their suggestions.

LEVEL I INSTRUMENT

The purpose of the Level I assessment was to sample the ability of candidates from the experimental and control groups to make appropriate decisions when designing or teaching a model-related lesson rather than to sample the domain of a candidate's knowledge about the Concept Attainment Model. For instance, a candidate was not asked to perform tasks such as recognizing or listing four phases of a model or recognizing definitions of terms. He/She was asked to determine what behaviors are appropriate in specified situations. A KR-20 reliability coefficient (Dick and Hagerty, 1971) of .645 was computed for the instrument.

LEVEL II INSTRUMENT

The Level II observation instrument consists of a list of twenty-one selected competencies and columns in which raters can indicate the tentative and confirmed presence of each competency. The competencies were selected from the results of an application of the Competency Derivation Procedure (Dodl and Freedman, 1974). The two three-person teams selected to use the instrument while observing videotapes practiced with two additional videotaped Concept Attainment lessons, and both teams independently agreed that exactly the same competencies were demonstrated on both tapes, thus establishing inter-rater reliability for the observation process.

LEVEL III INSTRUMENT

An attempt was made to identify all of the pupil classroom behaviors which were implied by the Concept Attainment Competency List, and an instrument was developed for recording the presence or absence of each behavior in a lesson. The procedure used to establish inter-rater reliability for the Level II observation process was equally successful for the Level III process.

LEVEL IV PRETEST

A Level IV pretest was designed to be administered by each teacher candidate to the pupils who were

taught by that candidate during the study. The test required pupils to attain two concepts, one involving geometric figures and another involving words. For each concept, each pupil was given seventeen cards containing examples and non-examples of instances of the concept.

Pupils were told that the teacher (candidate) had an idea and they should try to guess the idea. Pupils were also told that on a card with a YES typed on it was an example of the teacher's idea and a card with a NO typed on it did not include an example of the idea. If a pupil wanted to know if any of the other cards contained an example of the idea, she/he could hand the card to the teacher, and the teacher would write a YES or NO on the card and give it back to the pupil. However, the pupils were told that one of their objectives was to guess the idea while having as few cards as possible identified as containing examples or non-examples. When a pupil thought he/she had identified the concept, she/he wrote it on a blank card or drew a new example and gave it to the teacher. The teacher would write a YES or NO on the card and give it back.

A pupil demonstrated attainment of the concept by performing two tasks:

1. The concept was written on the blank card or a new example was drawn on a blank card.

2. All of the cards, including those identified by the teacher, were divided into stacks of examples and non-examples.

Pupils were given ten minutes to identify each concept. In the pretest, no pupils were able to attain either concept, so no reliability coefficient was computed, and scores were not included in computations for analysis.

LEVEL IV POSTTEST

As teacher candidates taught Concept Attainment lessons, it became apparent that many pupils were quickly mastering the skills measured on the pretest. Therefore, the following three steps were taken to increase the difficulty of the posttest.

1. Pupils were required to attain four concepts instead of two.
2. The concepts addressed were more difficult.
3. The pupils were allowed seven instead of ten minutes to attain each concept.

Use of a Kuder-Richardson formula for estimating reliability of essay tests (Ebel, 1972) resulted in a coefficient of .736.

RESULTS

At each level of assessment, correlation and multiple regression coefficients were computed from collapsed training-group scores. Since no pupils could perform any tasks on the Level IV pretest, only posttest scores were involved in computations.

CORRELATION COEFFICIENTS

Table 1 is a summary of correlation coefficients computed from scores at all levels of assessment. Correlation coefficients computed from scores at Levels I (teacher knowledge) and IV (pupil outcomes) were statistically significant at an alpha level of .05, $r\ (22df) = .552$, $p. < .01$; as was the coefficient computed from scores at Levels II (teacher classroom performance) and IV, $r\ (22df) = .447$, $p. < .05$. No other coefficients were statistically significant.

MULTIPLE REGRESSION

A multiple regression coefficient was computed from scores at all levels with Level IV as the dependent variable. Table 2 is a summary of the results of stepwise multiple regression techniques.

The first variable entered into computations was teacher knowledge (Level I). The resulting multiple R coefficient (.552) was statistically significant, $F\ (1, 22df) = 9\ .623$, $p. < .01$. The amount of variance explained by the linear relationship of Level I and Level IV scores, expressed as R square, was .304. Teacher

Table 1: Summary of correlation coefficients for scores at all levels of assessment*

Assessment Levels	I	II	III	IV
I	1.000	.249	.365	.552†
II		1.000	.356	.447††
III			1.000	.313
IV				1.000

*Level I: number of correct responses on cognitive-level multiple-choice test
Level II: number of model-related competencies demonstrated in a videotaped lesson
Level III: number of model-related pupil behaviors demonstrated in a videotaped lesson
Level IV: mean score of pupils on model-related pupil outcomes test
†p. $<$ .01
††p. $<$ .05

classroom performance (Level II) was the second variable entered. The resulting multiple R coefficient (.637) was statistically significant, F (2, 21df) = 7.188, p. $<$ .088. R square was .406, an increase of .102 from step 1. Pupil classroom behavior (Level III) was entered at step 3. Once again the resulting multiple R coefficient (.638) was statistically significant, F (3, 20df) = 4.578, p. $<$.026, but the increase in R square from step 2 was only .001.

Table 2: Summary of results of multiple regression

Step	1	2	3
Entered Variable	Teacher Knowledge (Level I)	Teacher Classroom Performance (Level II)	Pupil Classroom Behavior (Level III)
Multiple R	.552	.637	.638
R Square	.304	.406	.407
Change in R Square		.102	.001
F	9.623	7.188	4.578
P	.01	.008	.026

PREDICTIVE VALIDITY

Since scores from Assessment Levels I and II correlated significantly with pupil outcome scores, it appears that the teacher knowledge and teacher classroom performance assessment processes have value for predicting a teacher candidate's ability to positively influence pupil performance. If the assessment procedures provide valid measures of competencies related to teacher knowledge and teacher classroom performance, there is also an implication that those competencies are valid in terms of pupil outcomes. The more competencies demonstrated by a candidate, the better pupils taught by that candidate are likely to perform on assessments of concept attainment skills. However, that implication was not actually examined in the study, for no attempt was made to determine if candidates consistently used in lessons the competencies demonstrated in the videotaped lessons. That question can be easily addressed in future studies. There was no indication that Level III assessments have predictive validity or that competencies associated with inducing appropriate pupil classroom behaviors were effective for improving pupil concept attainment skills.

Since scores at Levels I and II correlated significantly with scores at Level IV, but did not correlate significantly with each other, a high multiple R coefficient for Level I and II scores with Level IV scores as a dependent variable would be expected. In fact, the amount of variance (.40) explained by a linear relationship between the scores is remarkably high for data collected in a field setting. Combinations of scores from Levels I and II would appear to have good predictive value. Although when Level III scores were added in stepwise regression the resulting multiple correlation coefficient was significant, the increase in the amount of variance explained by linear relationships was negligible. Thus it appears that Level III scores provide very little additional predictive value when combined with Level I and II scores.

CONCLUSION

Of course, this field test involved only one model of teaching, candidates in one PBTE program, and pupils in one grade; therefore, the generalizability of the study is limited. However, the potential of the system has been demonstrated, and at least two conclusions are implied in field test results:

1. Within a models of teaching framework, pupil outcome scores can be successfully used to estimate the predictive validity of lower-level assessment instruments.

2. Within a models of teaching framework, it may be possible to validate teaching competencies in terms of pupil outcomes.

Overall, the results of the study are encouraging. If replications with different models, grade levels, and programs are successful, the system should prove valuable for assessing competencies, designing effective instruments, improving programs, and perhaps even demonstrating the worth of the PBTE and models of teaching approaches. More important, however, is the potential of the system for providing a teacher behavior-pupil performance research model which generates successful results.

Obviously, if a models of teaching framework for competency assessment is to be feasible, it will also be necessary to design training programs involving the models of teaching approach. Such programs are currently being attempted (e.g., Columbia Teacher College), and research studies at Florida State University are being implemented in an effort to determine if such programs would encompass the domain of teaching skills needed by beginning teachers (Wheeler, 1978).

BIBLIOGRAPHY

Anastasi, A. *Psychological testing.* New York: Macmillan, 1968.

Berliner, D. C. *The acquisition of knowledge in the classroom.* Paper presented at the Conference on Schooling and the Acquisition of Knowledge, sponsored by the U. S. Navy Personnel Research and Development Center, San Diego, California, 1975.

Borg, W. R., Kelley, P. L., and Gall, M. D. *The minicourse: a microteaching approach to teacher education.* Beverly Hills, California: Macmillan Educational Services, 1970.

Burke, J. B., Hansen, J. H., Houston, W. R., and Johnson, C. *Criteria for describing and assessing competency based programs.* New York: The National Consortium of Competency Based Education Centers, 1975.

Cooper, J. M., Weber, W. A., and Johnson, C. E. *A systems approach to program design.* Berkeley, California: McCutchan, 1973.

Dick, W. and Hagerty, N. *Topics in measurement.* New York: McGraw-Hill, 1971.

Dodl, N. R. and Freedman, S. P. *A procedure for deriving competencies from models of teaching.* Tallahassee, Florida: Teacher Education Projects, 1974.

Ebel, R. L. *Essentials of educational measurement.* Englewood Cliffs, New Jersey: Prentice-Hall, 1972.

Ekstrom, R. B. The use of achievement tests to measure pupil growth. In D. A. Potter (Ed.), *A critical review of the literature: teacher performance and pupil growth.* Princeton, New Jersey: Educational Testing Service, 1973, 44-50.

Freedman, S. P. *An examination and validation of a procedure for deriving teacher competencies from models of teaching.* Unpublished doctoral dissertation, Florida State University, 1975.

Gagne, R. M. Policy implications and future research. In *Do teachers make a difference? A report on recent research on pupil achievement.* Washington, D. C.: Bureau of Educational Personnel Development, Office of Education, 1970.

Gall, M. D. *The problem of "student achievement" in research on teacher effects.* Paper presented at the national meeting of the American Educational Research Association, New Orleans, 1973.

Heath, R. W. and Neilson, M. A. The research basis for performance-based teacher education. *Review of Educational Research*, 1974, *44* (4), 463-484.

Hunt, D. E., Joyce, B. R., Greenwood, J., Noy, J. E., Reid, R., and Weil, M. *Student conceptual level and models of teaching: theoretical and empirical coordination of two models.* Paper presented at the national meeting of the American Educational Research Association, 1974.

James, B. N. An evaluation of an assessment system for performance-based teacher education programs (Doctoral dissertation, Florida State University, 1977). *Dissertation Abstracts International*, 1977, *38*, 2058A, No. 4. (University Microfilms No. 77-22119, 225)

Joyce, B. and Weil, M. *Models of teaching.* Englewood Cliffs, New Jersey: Prentice-Hall, 1972.

McDonald, F. J. Evaluation of teacher behavior. In W. R. Houston and R. B. Howsam (Eds.), *Competency-based teacher education: progress, problems, and prospects.* Chicago: Science Research Associates, 1972.

Medley, D. M., Soar, R. S., and Soar, R. *Assessment and research in teacher education: focus on PBTE.* Washington, D. C.: American Association of Colleges for Teacher Education, 1975.

Potter, D. A. *A research strategy for performance-based education.* Paper presented at the national meeting of the American Educational Research Association, New Orleans, 1973.

Rosenshine, B. and Furst, N. Research on teacher performance criteria. In B. O. Smith (Ed.), *Research in teacher education: a symposium.* Englewood Cliffs, New Jersey: Prentice-Hall, 1971.

Wheeler, A. M. *An examination of teaching competencies derived from models of teaching and their relationship to generic skills.* Unpublished doctoral dissertation, Florida State University, 1978.